

Atty. Docket No. YOR920010539US1  
(590.076)

**Amendments to the Claims:**

This listing of claims will replace all prior versions, and listings, of claims in the application:

**Listing of Claims:**

1. (Currently Amended) An apparatus for facilitating clustering of speech and audio data, said apparatus comprising:

an arrangement for obtaining speech and audio data as input data; and

an arrangement for creating a predetermined number of non-overlapping subsets of the input data;

said arrangement for creating a predetermined number of non-overlapping subsets being adapted to split the input data recursively;

said clustering being independent of any ~~system~~ model wherein the splitting of the input data into a predetermined number of non-overlapping subsets does not depend on a model.

2. (Original) The apparatus according to Claim 1, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to initially split the input data into at least two sets of output data.

3. (Original) The apparatus according to Claim 2, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to:

Atty. Docket No. YOR920010539US1  
(590.076)

split the at least two sets of output data recursively; and

repeat the recursive splitting of output data sets until the predetermined number of non-overlapping subsets is obtained.

4. **(Original)** The apparatus according to Claim 2, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine an eigenvector decomposition relating to the input data.

5. **(Original)** The apparatus according to Claim 4, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine a vector of projection coefficients onto the set of eigenvectors in the eigenvector decomposition.

6. **(Previously Presented)** The apparatus according to Claim 5, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine a probability distribution relating to the vector of projection coefficients.

7. **(Original)** The apparatus according to Claim 6, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to:

assign at least one threshold relating to the probability density; and

yield the at least two sets of output data based on the relation to the threshold of a value associated with a function relating to the projection coefficients.

Atty. Docket No. YOR920010539US1  
(590.076)

8. **(Original)** The apparatus according to Claim 7, wherein there are N-1 thresholds, where N is the number of sets of output data to be yielded.

9. **(Previously Presented)** The apparatus according to Claim 8, wherein each threshold is a value of the function relating to the projection coefficients for which the probability distribution equals  $m/N$ , where m is a number from 1 to N-1.

10. **(Original)** The apparatus according to Claim 1, wherein the data clustering relates to the enrollment of target speakers in a speaker verification system.

11. **(Currently Amended)** A method of facilitating clustering of speech and audio data, said method comprising the steps of:

obtaining speech and audio data as input data; and

creating a predetermined number of non-overlapping subsets of the input data;

step of creating a predetermined number of non-overlapping subsets comprising splitting the input data recursively;

said clustering being independent of any system model wherein the splitting of the input data into a predetermined number of non-overlapping subsets does not depend on a model.

12. **(Original)** The method according to Claim 11, wherein said splitting step comprises initially splitting the input data into at least two sets of output data.

Atty. Docket No. YOR920010539US1  
(590.076)

13. **(Original)** The method according to Claim 12, wherein said splitting step comprises:

splitting the at least two sets of output data recursively; and

repeating the recursive splitting of output data sets until the predetermined number of non-overlapping subsets is obtained.

14. **(Original)** The method according to Claim 12, wherein said splitting step comprises determining an eigenvector decomposition relating to the input data.

15. **(Original)** The method according to Claim 14, wherein said splitting step further comprises determining a vector of projection coefficients onto the set of eigenvectors in the eigenvector decomposition.

16. **(Previously Presented)** The method according to Claim 15, wherein said splitting step further comprises determining a probability distribution relating to the vector of projection coefficients.

17. **(Previously Presented)** The method according to Claim 16, wherein said splitting step further comprises:

assigning at least one threshold relating to the probability distribution; and

yielding the at least two sets of output data based on the relation to the threshold of a value associated with a function relating to the projection coefficients.

Atty. Docket No. YOR920010539US1  
(590.076)

18. **(Original)** The method according to Claim 17, wherein there are  $N-1$  thresholds, where  $N$  is the number of sets of output data to be yielded.

19. **(Previously Presented)** The method according to Claim 18, wherein each threshold is a value of the function relating to the projection coefficients for which the probability distribution equals  $m/N$ , where  $m$  is a number from 1 to  $N-1$ .

20. **(Original)** The method according to Claim 1, wherein the data clustering relates to the enrollment of target speakers in a speaker verification system.

21. **(Currently Amended)** A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for facilitating clustering of speech and audio data, said method comprising the steps of:

obtaining speech and audio data as input data; and

creating a predetermined number of non-overlapping subsets of the input data;

step of creating a predetermined number of non-overlapping subsets comprising splitting the input data recursively;

said clustering being independent of any ~~system model~~ wherein the splitting of the input data into a predetermined number of non-overlapping subsets does not depend on a model.